Available online at www.sciencedirect.com
ScienceDirect
Virology 373 (2008) 181–188
VIROLOGY
www.elsevier.com/locate/yviro

Protein coding content of the $U_L b'$ region of wild-type rhesus cytomegalovirus

Kristie L. Oxford, Meghan K. Eberhardt, Kai-Wen Yang, Lisa Strelow, Suzanne Kelly, Shan-Shan Zhou, Peter A. Barry *

Center for Comparative Medicine; Department of Pathology and Laboratory Medicine; California National Primate Research Center; University of California, Davis; Davis, CA 95616, USA

Received 18 September 2007; returned to author for revision 15 October 2007; accepted 31 October 2007
Available online 20 February 2008

Abstract

A recent comparison of two rhesus cytomegalovirus (RhCMV) genomes revealed that the region at the right end of the U_L genome component ($U_L b'$) undergoes genetic alterations similar to those observed in serially passaged human cytomegalovirus (HCMV). To determine the coding content of authentic wild-type RhCMV in this region, the $U_L b'$ sequence was amplified from virus obtained from naturally infected rhesus macaques without passage *in vitro*. A total of 24 open reading frames (ORFs) potentially encoding >99 amino acid residues were identified, 10 of which are related to HCMV ORFs and 15 to previously listed RhCMV ORFs. In addition, the analysis revealed a cluster of three novel alpha chemokine-like ORFs, bringing the number of predicted alpha chemokine genes in this region to six. Three of these six genes exhibit a high level of sequence diversity, as has been observed for the HCMV alpha chemokine gene *UL146*.

Keywords: Rhesus cytomegalovirus; Alpha chemokine; CXC chemokine; $U_L b'$; Sequence diversity; Human cytomegalovirus; *UL146*

Rhesus cytomegalovirus (RhCMV) and human cytomegalovirus (HCMV) are ubiquitous throughout their host populations (rhesus macaques and humans, respectively) and cause subclinical, persistent, lifelong infections in healthy individuals (Barry and Chang, 2007; Britt, 2007). In immune compromised individuals, however, CMVs can cause significant morbidity and mortality. RhCMV and HCMV genomes are largely colinear, although species-specific genes are present in each. The region at the right end of the U_L genome component ($U_L b'$) in HCMV contains multiple open reading frames (ORFs) encoding cell tropism and immune modulating functions that may be mutated or deleted during passage of clinical isolates *in vitro* (Cha et al., 1996; Dolan et al., 2004). Recent work has demonstrated that the corresponding region in RhCMV undergoes similar changes, thus the genetic content of $U_L b'$ in wild-type virus is presently undefined (Hansen et al., 2003; Rivailler et al., 2006).

The prototypical RhCMV strain 68-1 (Asher et al., 1974) has been annotated as containing 230 ORFs of >99 amino acid residues (aa) (Hansen et al., 2003) and the 180.92 strain as containing 258 ORFs ranging in size from 21 to 2178 aa (Rivailler et al., 2006). The analytical approaches taken were not the same in both studies, and the number and locations of genuine protein-coding ORFs remain unclear. There are also differences between the two strains, since the genome of 180.92 is approximately 5 kbp smaller than that of 68-1. The majority of the potential coding differences are located in a 15 kbp region that corresponds to $U_L b'$ in HCMV, and it appears that, as in HCMV, RhCMV $U_L b'$ is labile during passage *in vitro*. Comparisons of 68-1 and 180.92 have led to a model in which the sequence of wild-type RhCMV was proposed to contain four contiguous genomic segments designated A, B, C, and D (Fig. 1) (Rivailler et al., 2006). The sequence of 180.92 lacks segments C and D, while that of 68-1 lacks segment A, and segments B and C are inverted relative to segment D. In order to characterize the coding capacity of $U_L b'$ in wild-type RhCMV that has not been passaged *in vitro* and test this proposed model, DNA from RhCMV naturally circulating at the California National Primate

* Corresponding author. Center for Comparative Medicine, University of California, Davis, County Rd. 98 & Hutchison Dr., Davis, CA 95616, USA.
E-mail address: pabarry@ucdavis.edu (P.A. Barry).

0042-6822/$ - see front matter © 2007 Published by Elsevier Inc.
doi:10.1016/j.virol.2007.10.040

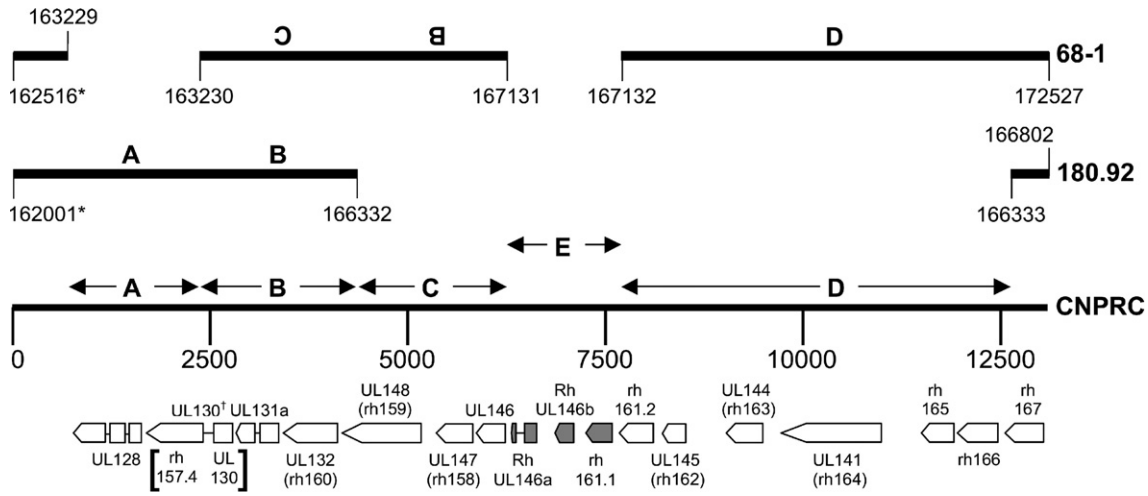

Fig. 1. Genomic arrangement of the $U_L b'$ region of wild-type RhCMV circulating in the rhesus macaque colony at the CNPRC (accession number EF990255) relative to the $U_L b'$ regions of RhCMV 68-1 (AY186194) and 180.92 (DQ120516). The regions of overlap with 68-1 and 180.92 are listed by the corresponding numbering system for each RhCMV variant. The proposed structure of the consensus RhCMV $U_L b'$ region (segments A – D) proposed by Rivailler et al. (2006) is shown together with the novel segment (E) found in RhCMV CNPRC. Predicted ORFs within $U_L b'$ region of RhMV CNPRC are presented, all of which are on the complementary strand. Shaded ORFs in segment E indicate those ORFs not present in either 68-1 or 180.92. ORFs homologous to HCMV ORFs are designated by 'UL'. ORFs without homologues in HCMV are designated with the naming scheme of Hansen et al (indicated by 'rh'). Boxes joined by a horizontal line represent proposed exons. rh161, annotated in 68-1, is listed here as rh161.2. rh161.1 is named to reflect the sequence identity with the previously named rh161 ORF (see text for details). RhUL146a and 146b are named to reflect their alpha chemokine-like sequence homology, similar to RhUL146. It should be noted that the 5′ point of overlap with 68-1 and 180.92 is imprecise (see text for details). †: The proposed RhUCMV L130 ORF based on a possible spliced product of the rh157.4 and RhCMV UL130 ORF (annotated in 180.92; see Fig. 3 and text for details).

Research Center (CNPRC) ($RhCMV_{CNPRC}$) was purified from seropositive animals, and the $U_L b'$ sequence of RhCMV was determined.

Results

A 13,066 bp sequence (46% G+C) of $RhCMV_{CNPRC}$ spanning the region of non-colinearity between the 68-1 and 180.92 genomes was assembled from multiple overlapping amplicons (GenBank Accession Number EF990255). The sequence included 714 and 470 bp of overlap with both genomes at its 5′ and 3′ ends, respectively (Fig. 1). The 5′ end of the sequence begins upstream of UL128, and the 3′ end of the sequence maps downstream of the rh167 ORF. The 5′ end of the sequence contains a 22 bp sequence (CCGTCTCTCAGAC-CAAATTTAC) directly repeated 30 times, with 27 copies identical in sequence. This sequence does not match any known transcription factor binding sites (data not shown). A variant sequence in the overlapping region of the 68-1 genome (CCGTCTCTCAGACTAATTTGAC) is present as 19 imperfect direct repeats, and another variant sequence (CCGGGTCTCA-GACCAATTTTAC) is present as 15 imperfect direct repeats in 180.92 (underlined nucleotides represent differences from $RhCMV_{CNPRC}$). Due to the number and degeneracy of the repeats in 68-1 and 180.92, there was some uncertainty about the exact extent of overlap with the 5′ end of the $RhCMV_{CNPRC}$ sequence.

Analysis of $RhCMV_{CNPRC}$ largely validated the model for the $U_L b'$ region (Rivailler et al., 2006). A total of 24 ORFs of >99 aa were identified in both strands, 15 of which (transcribed right to left, Fig. 1) are present in either or both of 68-1 and 180.92. Ten of these 15 ORFs have HCMV sequence homologues. Six of the identified ORFs, oriented from left to right (not shown in Fig. 1) were considered unlikely to represent authentic coding regions since they are not conserved in other primate cytomegalovirus genomes and exhibit no significant similarity to other proteins. The remaining 3 ORFs (RhUL146a, RhUL146b, and rh161.1 in segment E; Fig. 1) are absent from 68-1 and 180.92. Each of these novel genes potentially encodes a protein that is related (31–57% identity) to cellular alpha chemokines (Fig. 2).

Alpha chemokines contain a characteristic motif (CXC), and activation of target cells is dependent on the three residues (ELR) immediately preceding this motif (Clark-Lewis et al., 1991). Two of the novel $RhCMV_{CNPRC}$ ORFs (RhUL146b and rh161.1) contain this (ELRCXC) or a very similar motif (ELYCXC or EQRCXC, respectively). The third ORF (RhUL146a) contains another variant ([H/Y]PRCXC). RhUL146a is proposed to contain an intron since protein sequence homology to alpha chemokines is present in two discontinuous regions separated by appropriately located splice sites (Fig. 2C). The proposed protein product of this gene exhibits strong structural homology to alpha chemokines based on Phyre analysis. Although BLAST analysis did not identify any viral protein sequences homologous to RhUL146a, CCMV but not HCMV also contains an ORF, UL146A, in the same general region within $U_L b'$ (Davison et al., 2003). CCMV UL146A is related to the tandemly arrayed UL146 ORF, which is homologous to HCMV UL146, an alpha chemokine (Penfold et al., 1999).

The presence of the three novel RhCMV genes and the genomic arrangement of $U_L b'$ were confirmed using viral DNA amplified from two additional non-co-housed macaques

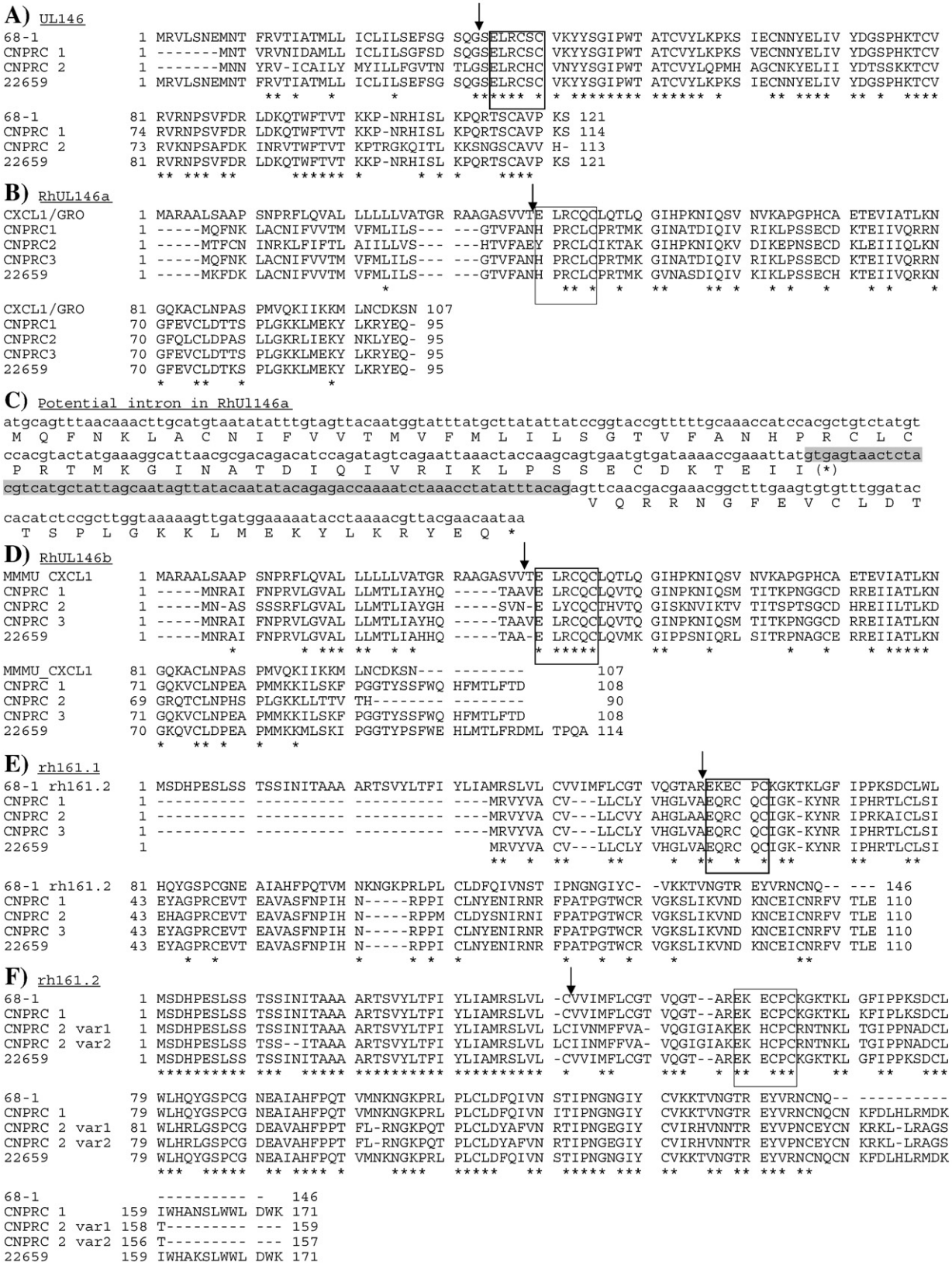

Fig. 2. Protein alignments of (A) RhUL146, (B) RhUL146a, (D) RhUL146b, (E) rh161.1, and (F) rh161.2 of $RhCMV_{68\text{-}1}$, $RhCMV_{CNPRC}$, and $RhCMV_{22659}$. $RhCMV_{CNPRC}$ protein sequences isolated from different macaques are indicated as CNPRC1, CNPRC2, or CNPRC3. If two variants were isolated from the same macaque, they are designated as var1 and var2. The proposed exon/intron boundaries for RhUL146a are also shown (C). Predicted signal peptide cleavage sites are illustrated by an arrow; the arrow in (E) refers to the predicted signal peptide cleavage site for rh161.1. Boxed amino acids represent the ELRCXC-like motif. Amino acids conserved in all of the aligned proteins are indicated by an asterisk. RhUL146a is aligned with the chemokine ligand 1 (CXCL1/GRO) of rhesus macaque (*Macaca mulatta*) (accession number NP_001028050).

(accession numbers EF990256 and EU003822) and a clinical RhCMV isolate (22659) (accession number EU130540) that had undergone a limited but unknown number of passages in culture (Alcendor et al., 1993; Barry et al., 1996). Therefore, the U_Lb′ structure presented in Fig. 1 likely reflects that of wild-type RhCMV.

Flanking these novel genes in segment E are three other alpha chemokine-like ORFs found only in the 68-1 RhCMV strain: *UL146* (accession number AAO40076, not annotated in 68-1; Penfold, personal communication), *UL147* (annotated as rh158 in 68-1), and *rh161.2* (Fig. 1). ORF rh161.2, originally described in 68-1 as rh161, shares 39% identity with rh161.1 (Fig. 2E). The conserved regions of rh161.1 and 161.2 include a consensus alpha chemokine motif (EQRCQC in rh161.1 and EKE/HCPC in rh161.2), in addition to 7 other cysteine residues. It is likely that rh161.1 and rh161.2 arose via a duplication event, and the nomenclature assigned herein reflects this. Although sequence identity of rh161.2 with alpha chemokines is limited, Phyre analysis revealed some structural homology to interleukin-8 around the CXC motif. While functional studies remain to be performed, it appears that wild-type RhCMV encodes up to six alpha chemokine-like proteins in the U_Lb′ region.

Comparison of wild-type RhCMV U_Lb′ sequences with those from passaged strains revealed a high sequence conservation in most ORFs (Table 1), mirroring what is generally found between HCMV clinical and tissue culture-passaged isolates. Exceptions to this in RhCMV appear to be limited to the alpha chemokine-like genes. *UL146* in wild-type variants were as much as 43% divergent from *UL146* in 68-1. *UL146* in HCMV clinical isolates can also diverge by as much as 60% (Arav-Boger et al., 2006; Hassan-Walker et al., 2004; Lurain et al., 2006; Prichard et al., 2001). Rh161.2 in wild-type RhCMV was also highly variable from that in 68-1 (72–100% aa identity) in addition to differing in length by 12 to 25 aa at the carboxyl terminus (Fig. 2F). One of the breakpoints for DNA sequence discontinuity between $RhCMV_{68\text{-}1}$ and $RhCMV_{CNPRC}$ (nucleotide 167312 in 68-1, Fig. 1) is immediately downstream of the reported stop codon for rh161 in 68-1. Therefore, it appears that the rearrangement of the 68-1 genome eliminated the carboxy terminal 12–25 aa of this particular ORF, and generated a novel

Table 1
ORFs in the U_Lb′ region of $RhCMV_{CNPRC}$

Strand[a]	RhCMV ORF	HCMV ORF	From	To	Length (aa)	Exon	ID with RhCMV	ID with HCMV	Comment
C	None	None	1	747	249				
	None	None	2	793	264				
	None	None	3	761	253				
	None	None	9459	9812	118				
	None	None	11411	11848	146				
	None	None	12311	12619	102				
W	RhUL128	UL128	722	1175	150	3	100	44	Homology between aa 15 and 134 of HCMV UL128
			1261	1398	46	2			
			1487	1626	47	1			
	rh157.4	NH	1628	2359	243		100		14 aa longer than 180.92 rh157.4
	RhUL130	UL130	2481	2762	93		97	44	Homology between aa 56 and 114 of HCMV UL130
	RhUL131a	UL131A	2759	3020	86	2	99	31	
			3103	3359	86	1			
	RhUL132 (rh160)	UL132	3403	4068	221		99	31	
	RhUL148 (rh159)	UL148	4134	5114	326		99	30	Homology between aa 14 and 264 of HCMV UL148
	RhUL147 (rh158)	UL147	5319	5780	153		96	36	CXC motif
	RhUL146	UL146	5831	6175	114		92/57/100*	27/33/27	CXC motif;
	RhUL146a[E]	NH	6252	6348	31	2			CXC motif
			6425	6615	63	1			
	RhUL146b[E]	NH	6738	7064	108				CXC motif
	rh161.1[E]	NH	7188	7520	110				CXC motif; possible tandem duplication of rh161.2 (see text)
	rh161.2	NH	7602	8117	157–171		72–100		CXC motif; annotated as rh161 in 68–1; 12–25 aa longer than rh161 of 68-1; see Fig. 2A and text
	RhUL145 (rh162)	UL145	8202	8507	101		100	65	
	RhUL144 (rh163)	UL144	8925	9440	171		98	31	
	RhUL141 (rh164)	UL141	9643	10935	430		99	44	
	rh165	NH	11396	11827	143		100		Possible tandem duplication of rh166
	rh166	NH	11872	12396	174		100		Possible tandem duplication of rh166
	rh167	NH	12527	13030	167		98		Possible tandem duplication of rh165

NH: no homologue in HCMV.
[E]: ORFs found in segment E of $RhCMV_{CNPRC}$ and $RhCMV_{22659}$ (see Fig. 1).
*: CNPRC 1/CNPRC 2/22659 (see Fig. 2A).
[a] C = coding strand (left-to-right transcription); W = complementary strand (right-to-left transcription).

stop codon in the process. Two partial UL147 clones also exhibited considerable sequence divergence in the amino terminus (54%) (data not shown), similar to what has been reported for HCMV UL147 (Arav-Boger et al., 2006; Lurain et al., 2006). Further analysis is required to confirm this interpretation. High sequence divergence was similarly prevalent in the 3 novel ORFs in segment E. Wild-type sequences diverged from each other by 10% for rh161.1, 44% for RhUL146b, and 46% for RhUL146a (Figs. 2E, D, and B). RhUL146b was also distinguished by differences in the length of the predicted protein.

The tandemly arrayed rh165 and rh166 ORFs appear to have arisen from a single ancestral gene because the wild-type RhCMV sequences exhibit 39% identity with each other. Both wild-type rh165 and 166 are 100% identical to their 68-1 homologues. Phyre analysis does not indicate structural homology with any viral or cellular proteins. The annotation for $RhCMV_{68\text{-}1}$ rh166 indicates that the sequence is "similar to UL133" of HCMV (Hansen et al., 2003), however the wild-type rh166 does not share any evident sequence homology with HCMV UL133. Only a weak homology to chimp CMV (CCMV) UL138 (36% identity to aa 25–50 of NP_612763) is present in rh166 (data not shown).

Wild-type RhCMV rh157.4 (no homologue in HCMV) and RhUL130 are both homologous to HCMV UL130 within the carboxy terminal two-thirds and the amino terminal one-third of the two ORFs, respectively. The wild-type rh157.4 encodes an additional 14 aa at the carboxyl terminus that are not present in the rh157.4 ORF of $RhCMV_{180.92}$. One of the rh157.4 clones also contains an internal 16 aa deletion. Apart from these changes, rh157.4 of $RhCMV_{180.92}$ and wild-type RhCMV are

A) Proposed Exon/Intron boundaries for RhUL130/rh157.4

```
atgaaaaactataatcaacaactctatacctttgttttgctgatgctattgcgttatggct
 M  K  N  Y  N  Q  Q  L  Y  T  F  V  L  L  M  L  L  R  Y  G   20
ttcggtgatacttcttctatcaaggatcttaaactcggcaatttttgtcccatgtattca
 F  G  D  T  S  S  I  K  D  L  K  L  G  N  F  C  P  M  Y  S   40
tctcctcgacgaaattttgggttgtttacatcttacaaaactcatccaacgggaccggaa
 S  P  R  R  N  F  G  L  F  T  S  Y  K  T  H  P  T  G  P  E   60
tgtggaaacaccggcctttacgtttttcacaataggtataatcaataccttattgaaaga
 C  G  N  T  G  L  Y  V  F  H  N  R  Y  N  Q  Y  L  I  E  R   80
ccatccgcgtggtcaaacaaactggcgttctacctcaggtgagttttgtatcataacgtat
 P  S  A  W  S  N  K  L  A  F  Y  L  S  (*)                    93
aatattcataatacagattataccaatacttttgggtgaacatatatttttataaatcat

tgtagtgcactacactcgcccgtgttccaaaagttttcgaaaatggctacatcgtctact
       A  L  H  S  P  V  F  Q  K  F  S  K  M  A  T  S  S  T   111
acatctatgaatattacggaagaagagaagaagactttcgctgcgcatatgatacccatg
 T  S  M  N  I  T  E  E  E  K  K  T  F  A  A  H  M  I  P  M   131
cgtagtacaactcttcgttacatagttaaggacggtactgacgtggaacactgtcagatg
 R  S  T  T  L  R  Y  I  V  K  D  G  T  D  V  E  H  C  Q  M   151
cgtgtcatcacttgggccaagacagaagctaatttcatctctttcaaagttaaaatagaa
 R  V  I  T  W  A  K  T  E  A  N  F  I  S  F  K  V  K  I  E   171
ctttcaaacgcttacagaagacctagtagcatttgtaccaggcctaatctagttgtccct
 L  S  N  A  Y  R  R  P  S  S  I  C  T  R  P  N  L  V  V  P   191
gaacctgataaaataagcccgacagtaaaaacaccaacagtaacaactcccaagcgaacg
 E  P  D  K  I  S  P  T  V  K  T  P  T  V  T  T  P  K  R  T   210
gcgagatcaacggtaaaagcaacgccaaggtcagctccagcaaagcaagtggcaccgaag
 A  R  S  T  V  K  A  T  P  R  S  A  P  A  K  Q  V  A  P  K   230
ccagcaacagcggcaaaacaggtagcagcgaaaccagctccagcaaagccagcaacagcg
 P  A  T  A  A  K  Q  V  A  A  K  P  A  P  A  K  P  A  T  A   250
gcaaagccggtagcagccaaaccagcggcaacaaagccagcgactaaaccagcagccgca
 A  K  P  V  A  A  K  P  A  A  T  K  P  A  T  K  P  A  A  A   270
aaacctgtagcggtaaagccagcggcagcaaaaccaccaacgtcggcagcaaaatcacaa
 K  P  V  A  V  K  P  A  A  A  K  P  P  T  S  A  A  K  S  Q   290
aacagtgcgcccaaaggccagcgaagccaactgtcacaaaagccaaaacctaaacccaaa
 N  S  A  P  K  G  Q  R  S  Q  L  S  Q  K  P  K  P  K  P  K   310
acaccagtaccaattaggtataaacaatacttccaatccaacggaaaaagaaggtacccg
 T  P  V  P  I  R  Y  K  Q  Y  F  Q  S  N  G  K  R  R  Y  P   330
ctagcgtcacgatatagttccgcctggctgtttaggcggcatccttccggctaatatgtt
 L  A  S  R  Y  S  S  A  W  L  F  R  R  H  P  S  G  *         347
```

B) Alignment of RhU:130/157.4 with CCMV UL130

```
RhUL130      35  FCP-MYSSPRRNFGLFTSYKTHPTGPECGNTGLYVFHNRYNQYLIERPSAWSNKLAFYLS  93
                 +CP+ + SP R    FT Y   P GP C N  L + +NR  Q LIERPS W  K+ +YLS
CCMV UL130   84  YCPLTFPSPPRRLSEFTRYHEVPGGPSCNNETLVLLYNRDGQTLIERPSPWVKKVTWYLS  143

RhUL130      94  ALHSPVFQKFSKMATS-STTSMNITEEEKKTFAAHMIPMRSTTLRYIVKDGTDVEHCQMR  152
                     PVFQ+F+K A T++  ++ I  ++ K F +HM+P ++  LR+++ DGT  E C M+
CCMV UL130 144  GRDHPVFQRFAKTASTANDGNVQIRTDDAKVFGSHMVPKQTKLLRFVINDGTRYEMCIMK  203

RhUL130     153  VITWAKTEANF-ISFKVKIELSNAYRRPSSICTRPNLVV  190
                 + TWA    ++ ++F+ ++  + A  +  S CT PNL+V
CCMV UL130 204  LETWAHVFRDYSVAFQTRLTFTQANNQTYSYCTHPNLLV  242
```

Fig. 3. Proposed splicing of the RhUL130 and rh157.4 ORF. (A) A putative intron for a spliced RhUL130 and rh157.4 fusion is indicated by shading. The first five amino acids of the unspliced rh157.4 ORF are underlined. *: Translation stop codon. (B) Alignment of the RhUL130/rh157.2 fusion with chimp CMV UL130 (CCMV; accession number NP_612749) with the site of the proposed exon/exon boundary indicated by the arrow.

100% identical. Phyre analysis does not reveal any significant homologies with viral or cellular proteins. Potential splice donor and acceptor sites were identified (GT and AG, respectively), which, if utilized, would generate a fusion of rh157.4 and RhUL130 (Fig. 3). The splice acceptor site is located 13 aa before the putative ATG codon of 157.4, seven of which are identical to CCMV UL130. Utilization of these splice sites, as well as those proposed for RhUL128, RhUL131a, and RhUL146a, remains to be confirmed.

Discussion

This report redefines the coding potential of the $U_L b'$ region of RhCMV by the identification of three novel alpha chemokine-like ORFs within multiple RhCMV variants that have never been passed in culture. The genomic structure of $U_L b'$ is conserved between RhCMVs currently circulating amongst CNPRC macaques and a viral isolate (22659) first passed in culture almost 20 years ago. This evidence strengthens the interpretation that the order of genes in Fig. 1 accurately represents the coding content of $U_L b'$ in wild-type RhCMV. The $U_L b'$ region of RhCMV, like the homologous region in HCMV, is uniquely unstable and can accumulate mutations and deletions even after brief passage *in vitro*. The basis for the genomic instability within RhCMV $U_L b'$ remains unknown. There are no obvious repeat or secondary structures, or distinct changes in base composition that might account for the high propensity to undergo rearrangements. Given the extent of the gene loss in many HCMV and RhCMV strains after passage in culture, it is clear that these ORFs are not required for replication in cultured fibroblasts. However, the absence of $U_L b'$ rearrangements in RhCMV 22659, which was propagated initially on MRC-5, human diploid fibroblasts, indicates that limited growth on non-host cells does not necessarily lead to genomic rearrangements. Reported *in vitro* functions for HCMV $U_L b'$-encoded ORFs strongly indicate that their functions contribute to cell tropism and viral pathogenesis and/or escape from host viral clearance mechanisms during establishment and maintenance of persistence. Thus it would appear that growth on cell types susceptible to productive infection *in vivo* should be used to propagate wild-type virus *in vitro*.

CMV species co-evolved with their cognate hosts during the radiation and speciation of mammals resulting in similar but not identical repertoires of viral immune modulating genes. Co-evolution of RhCMV with its macaque host appears to have selected for a heavy viral investment in alpha chemokine-like proteins. The presence of six ORFs in RhCMV with homology to alpha chemokines, versus only two in HCMV and three in CCMV, suggest that each virus has distinct evasion strategies against host antiviral mechanisms. Although the six RhCMV ORFs share sequence motifs and or structural characteristics with alpha chemokines, *in vitro* and *in vivo* studies are needed to determine the roles of these ORFs in RhCMV natural history. Depending on which amino acids flank the CXC motif, chemokine engagement of its receptor can be agonistic or antagonistic for signaling (Loetscher and Clark-Lewis, 2001; Moser et al., 1993). The promiscuity of chemokine/receptor interactions allows agonists of certain immune cells, such as T_H1 cells, to be antagonists of others, such as T_H2 cells. The presence of these ORFs in RhCMV may enable a subtle and highly refined modulation of host cell trafficking and activation during the earliest stage of RhCMV infection.

It is especially intriguing that these same ORFs are hypervariable in circulating strains of both RhCMV and HCMV. Both RhCMV and HCMV UL146 exhibit considerable sequence divergence, yet for HCMV UL146, functionality is retained despite the divergence (Prichard et al., 2001). HCMV UL146 binds to IL-8 receptors and induces neutrophil chemotaxis, calcium flux, and degranulation (Penfold et al., 1999), potentially facilitating viral dissemination. All three wild-type RhCMV UL146 variants maintained the ELRCXC motif required for chemokine receptor binding and activity (Clark-Lewis et al., 1991) and four other cysteine residues, implying preservation of tertiary structure. In addition to UL146, $RhCMV_{CNPRC}$ also encodes the HCMV UL147 homologue (rh158) and contains the conserved CXC motif. However, HCMV UL147 does not bind to the CXC receptor, and no chemokine-like function has been observed (Saederup and Mocarski, 2002). It should be noted that there is no rapid genetic drift of HCMV UL146 and 147 sequences *in vivo*. These sequences remain invariant when HCMV virus is repeatedly isolated from individual patients over time (Lurain et al., 2006). Partial clones suggest a similarly high amount of sequence divergence within RhCMV UL147. Together with the high variability, RhUL146a, RhUL146b, rh161.2, RhUL146 and RhUL147 are linked by an apparent involvement in targeting a similar pathway(s). Why the putative viral-mediated modulation of innate cell trafficking and signaling would be associated with profound sequence diversity remains a mystery.

Sequence analysis of cellular proteins has shown that host defense ligands and their receptors undergo the most rapid divergence compared to other cellular protein families. In addition, this diversity may arise from species-specific molecular mimicry by microbes (Murphy, 1993). Thus, it may be hypothesized that diversity found in both host immune defense proteins and viral immune modulatory proteins arises from a stepwise competition for fitness by each organism, where viral hijacking of cellular genes is a common theme.

Additional chemokine-like ORFs in wild-type RhCMV implies that viral modulation of innate immune responses is an especially critical mechanism during the earliest stage of virus infection. Evidence for the presence of multiple chemokines from this and other studies strengthens the hypothesis that RhCMV alters numerous host antiviral responses that cumulatively and ultimately favor viral dissemination and lifelong persistence. Additional studies are required to elucidate specific pathways targeted by these immunomodulatory proteins to subvert host immunity and favor viral persistence in an immune competent host.

Sequence homologues of HCMV and CCMV $U_L b'$ ORFs that were not identified in any RhCMV isolate (CNPRC, 22659, 68-1, and 180.92) include UL133, 135, 136, 139, 140, 142, 147a, 150, 151, and possibly 138. However, functional homologues may be encoded by the numerous ORFs found

only in RhCMV (rh157.4, RhUL146a, RhUL146b, rh161.1, rh161.2, rh165-167). Together with an absence of UL18, the lack of a UL142 homologue means that RhCMV does not appear to encode any of the MHC type I homologues found in HCMV. RhCMV does encode other ORFs implicated in HCMV attenuation of natural killer cell function, including UL40, 141, and a duplication of UL83, although functional studies have not yet been performed. Another HCMV ORF missing from RhCMV, UL138, has recently been implicated in the establishment and maintenance of latency in $CD34^+$ myeloid progenitor cells. It is not known at this time whether RhCMV infects the $CD34^+$ cell lineage in rhesus macaques and thus whether it encodes a functional homologue to HCMV UL138 or an analogous ORF that promotes latency in a different cellular reservoir. The presence of these additional ORFs in both CCMV and HCMV strongly implies that they conferred some selective advantage to the great ape CMVs in their respective hosts.

In future studies of the rhesus macaque/RhCMV model, the use of clinical strains of RhCMV possessing the entire $U_L b'$ region will enhance the understanding of host–virus dynamics involving tropism, latency, and pathogenesis. Moreover, a thorough comparison between the entire genomes of wild-type and laboratory passaged RhCMV may reveal potentially deleted non-core ORFs in regions other than $U_L b'$.

Materials and methods

Oral swabs were collected from RhCMV-seropositive rhesus macaques by running a Dacron swab along the gumlines and buccal pouch and allowing the swab to absorb a saturating amount of saliva (~0.2 ml). Viral DNA was purified from the unclarified swab samples according to previously published methods (Huff et al., 2003). The $U_L b'$ region of RhCMV was amplified in overlapping segments by PCR (*Taq* or *Pfu*) using primers corresponding to the sequences published for RhCMV strains 68-1 and 180.92 (accession numbers AY186194 and DQ120516, respectively) (Hansen et al., 2003; Rivailler et al., 2006). The amplicons were cloned using the pCR2.1 TA cloning vector (Invitrogen Corporation, Carlsbad, CA). Every ORF in Fig. 1 was analyzed with amplicons from at least two different animals and a clinical isolate, 22659 (Alcendor et al., 1993; Barry et al., 1996). Multiple independent clones from each of three animals were sequenced for the segment E genes. DNA sequences from overlapping clones were aligned using CHAOS-DIALIGN (Brudno et al., 2003), translation products were located with the ExPASy proteomics server (Gasteiger et al., 2003), protein sequences were aligned using the ClustalW multiple sequence alignment program (www.ebi.ac.uk/clustalw/# (Thompson et al., 1994)), relationships to other proteins were identified using the NCBI BLAST programs (www.ncbi.nlm.nih.gov/BLAST/), signal peptide cleavage sites were identified using SignalP 3.0 (www.cbs.dtu.dk/services/SignalP/ (Emanuelsson et al., 2007)), and the Protein Homology/Analogy Recognition Engine (Phyre) was used to analyze proteins for structural similarities to other proteins (Kelley et al., 2000).

The accession numbers for the RhCMV sequences are EF990255, EF990256, EU003822, and EU130540.

Note added in proof

The proposed splice site for the RhCMV UL130/rh157.4 ORFs (Fig. 3) has been confirmed by reverse transcription/PCR analysis using RNA isolated from cells infected with RhCMV strain 22659.

Acknowledgments

This work was supported by funding from the National Institutes of Health to PAB (AI063356 and AI49342-06) and the California National Primate Research Center (RR000169). We are grateful to Andrew Davison for comments on the manuscript.

References

Alcendor, D.J., Barry, P.A., Pratt-Lowe, E., Luciw, P.A., 1993. Analysis of the rhesus cytomegalovirus immediate-early gene promoter. Virol 194, 815–821.

Arav-Boger, R., Foster, C.B., Zong, J.C., Pass, R.F., 2006. Human cytomegalovirus-encoded alpha-chemokines exhibit high sequence variability in congenitally infected newborns. J. Infect. Dis. 193, 788–791.

Asher, D.M., Gibbs, J.C.J., Lang, D.J., Gadjusek, D.C., Chanock, R.M., 1974. Persistent shedding of cytomegalovirus in the urine of healthy rhesus monkeys. Proc. Soc. Exp. Biol. Med. 145, 794–801.

Barry, P.A., Chang, W.-L.W., 2007. Primate Betaherpesviruses. In: Arvin, A., Campadielli, G., Moore, P., Mocarski, E., Roizman, B., Whitley, R., Yamanishi, K. (Eds.), Human Herpesviruses: Biology, Therapy and Immunoprophylaxis. Cambridge University Press, Cambridge, pp. 1051–1075.

Barry, P.A., Alcendor, D.J., Power, M.D., Kerr, H., Luciw, P.A., 1996. Nucleotide sequence and molecular analysis of the rhesus cytomegalovirus immediate-early gene and the UL121117 open reading frames. Virology 215, 61–72.

Britt, W., 2007. Virus entry into host, establishment of infection, spread in host, mechanisms of tissue damage. In: Arvin, A., Campadielli, G., Moore, P., Mocarski, E., Roizman, B., Whitley, R., Yamanishi, K. (Eds.), Human Herpesviruses: Biology, Therapy and Immunoprophylaxis. Cambridge University Press, Cambridge, pp. 737–764.

Brudno, M., Chapman, M., Gottgens, B., Batzoglou, S., Morgenstern, B., 2003. Fast and sensitive multiple alignment of large genomic sequences. BMC Bioinformatics 4, 66.

Cha, T., Tom, E., Kemble, G., Duke, G., Mocarski, E., Spaete, R., 1996. Human cytomegalovirus clinical isolates carry at least 19 genes not found in laboratory strains. J. Virol. 70, 78–83.

Clark-Lewis, I., Schumacher, C., Baggiolini, M., Moser, B., 1991. Structure–activity relationships of interleukin-8 determined using chemically synthesized analogs. Critical role of NH_2-terminal residues and evidence for uncoupling of neutrophil chemotaxis, exocytosis, and receptor binding activities. J. Biol. Chem. 266, 23128–23134.

Davison, A.J., Dolan, A., Akter, P., Addison, C., Dargan, D.J., Alcendor, D.J., McGeoch, D.J., Hayward, G.S., 2003. The human cytomegalovirus genome revisited: comparison with the chimpanzee cytomegalovirus genome. J. Gen. Virol. 84, 17–28.

Dolan, A., Cunningham, C., Hector, R.D., Hassan-Walker, A.F., Lee, L., Addison, C., Dargan, D.J., McGeoch, D.J., Gatherer, D., Emery, V.C., Griffiths, P.D., Sinzger, C., McSharry, B.P., Wilkinson, G.W., Davison, A.J., 2004. Genetic content of wild-type human cytomegalovirus. J. Gen. Virol. 85, 1301–1312.

Emanuelsson, O., Brunak, S., von Heijne, G., Nielsen, H., 2007. Locating proteins in the cell using TargetP, SignalP and related tools. Nat. Protoc. 2, 953–971.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D., Bairoch, A., 2003. ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 31, 3784–3788.

Hansen, S.G., Strelow, L.I., Franchi, D.C., Anders, D.G., Wong, S.W., 2003. Complete sequence and genomic analysis of rhesus cytomegalovirus. J. Virol. 77, 6620–6636.

Hassan-Walker, A.F., Okwuadi, S., Lee, L., Griffiths, P.D., Emery, V.C., 2004. Sequence variability of the alpha-chemokine UL146 from clinical strains of human cytomegalovirus. J. Med. Virol. 74, 573–579.

Huff, J.E., Eberle, R., Capitanio, J., Zhou, S.-S., Barry, P.A., 2003. Differential detection of B virus and rhesus cytomegalovirus in rhesus macaques. J. Gen. Virol. 84, 83–92.

Kelley, L.A., MacCallum, R.M., Sternberg, M.J., 2000. Enhanced genome annotation using structural profiles in the program 3D-PSSM. J. Mol. Biol. 299, 499–520.

Loetscher, P., Clark-Lewis, I., 2001. Agonistic and antagonistic activities of chemokines. J. Leukoc. Biol. 69, 881–884.

Lurain, N.S., Fox, A.M., Lichy, H.M., Bhorade, S.M., Ware, C.F., Huang, D.D., Kwan, S.P., Garrity, E.R., Chou, S., 2006. Analysis of the human cytomegalovirus genomic region from UL146 through UL147A reveals sequence hypervariability, genotypic stability, and overlapping transcripts. Virol. J. 3, 4.

Moser, B., Dewald, B., Barella, L., Schumacher, C., Baggiolini, M., Clark-Lewis, I., 1993. Interleukin-8 antagonists generated by N-terminal modification. J. Biol. Chem. 268, 7125–7128.

Murphy, P.M., 1993. Molecular mimicry and the generation of host defense protein diversity. Cell 72, 823–826.

Penfold, M.E.T., Dairaghi, D.J., Duke, G.M., Saederup, N., Mocarski, E.S., Kemble, G.W., Schall, T.J., 1999. Cytomegalovirus encodes a potent α chemokine. Proc. Natl. Acad. Sci. U. S. A. 96, 9839–9844.

Prichard, M.N., Penfold, M.E., Duke, G.M., Spaete, R.R., Kemble, G.W., 2001. A review of genetic differences between limited and extensively passaged human cytomegalovirus strains. Rev. Med. Virol. 11, 191–200.

Rivailler, P., Kaur, A., Johnson, R.P., Wang, F., 2006. Genomic sequence of rhesus cytomegalovirus 180.92: insights into the coding potential of rhesus cytomegalovirus. J. Virol. 80, 4179–4182.

Saederup, N., Mocarski Jr., E.S., 2002. Fatal attraction: cytomegalovirus-encoded chemokine homologs. Curr. Top. Microbiol. Immunol. 269, 235–256.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.